

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

Introduction

This documentation is for the 2000 birth cohort linked birth/infant death data set (linked file). Previous birth cohort linked files were released for data years 1983-91. Beginning with 1995 data, the linked file was released in two different formats — period data and birth cohort data.

Period data — The numerator for the 2000 period linked file consists of all infant deaths occurring in 2000 linked to their corresponding birth certificates, whether the birth occurred in 2000 or 1999. The denominator file for this data set is the 2000 natality file, that is, all births occurring in 2000.

Birth cohort data — The numerator of the 2000 birth cohort linked file consists of deaths to infants born in 2000 linked to their corresponding birth certificates, whether the death occurred in 2000 or 2001. The denominator file is the 2000 natality file, that is, all births occurring in 2000.

The release of linked file data in two different formats allows NCHS to meet customer demands for more timely linked file data while still meeting the needs of data users who prefer the birth cohort format. For most general purposes, differences between the birth cohort and period linked files are negligible. However, birth cohort files are preferred for multivariate and some other types of detailed analysis because they follow a given cohort of births for an entire year to ascertain their mortality experience. This is generally considered to be a more robust methodology than the period file, which is essentially cross-sectional in nature.

The 2000 birth cohort linked file includes several separate data files. The first file includes linked birth and death certificate data for all US infants born in 2000 who died before their first birthday - referred to as the numerator file. The second file contains information from the death certificate for all US infant death records which could not be linked to their corresponding birth certificates - referred to as the unlinked death file. The third file is the 2000 NCHS natality file for the US with a few minor modifications - referred to as the denominator-plus file. These same three data files are also available for Puerto Rico, the Virgin Islands, and Guam.

For the denominator-plus file, selected variables from the numerator file have been added to the denominator file to facilitate processing. These variables include age at death (and recodes), underlying cause of death (and the 130-cause recode), place of accident, and record weight. These variables are the most widely used variables from the numerator file. With the previous file format it was sometimes necessary to combine the numerator and denominator files when performing certain multivariate statistical techniques. Now, when the number of variables required from the numerator file is limited, the denominator-plus file may be used by itself for ease of programming. Infant death identification numbers are also included, so that the same

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

infant can be uniquely identified and matched between the numerator and denominator-plus files.

Weighting

In part to correct for known biases in the data, changes were made to the linked file beginning with the 1995 data year. These changes include the addition of a record weight and an imputation for not-stated birthweight. In the 2000 birth cohort linked file, 98.7% of infant death records were linked to their corresponding birth certificates. Overall, 1.3% of infant death records could not be linked because the matching birth certificate could not be found; however this percent varied considerably by State (see section Table 1 below). Beginning with 1995 data, a record weight was added to the infant death records to correct in part for biases in percent of records linked by major characteristics. The number of infant deaths in the linked file are weighted to equal the sum of the linked plus unlinked infant deaths by age at death and state. The formula for computing the weights is as follows:

$$\frac{\text{number of linked infant deaths} + \text{number of unlinked infant deaths}}{\text{number of linked infant deaths}}$$

A separate weight is computed for each State of residence of birth and each age at death category (<1 day, 1-27 days, 28 days-1 year). Thus, weights are 1.0 for states which link all of their infant deaths. These weights have been added to all linked infant death records in the numerator file, and in the denominator-plus file. In the denominator-plus file, records for surviving infants have been assigned a weight of 1.0. This causes the denominator-plus file to weight up to about 358 (by residence) or 360 (by occurrence) more than the total number of live births (about 4 million), thus most runs on live birth data from the denominator-plus file should be run unweighted. Weights have not been computed for the Puerto Rico, Virgin Islands, and Guam files.

The addition of weighting to the file has greatly reduced bias, but has also created challenges for data analysis. The researcher should be aware that the use of the weights is appropriate for some, but not all applications. Weights should be used when computing the total number of infant deaths, or the number of infant deaths by characteristics, either from the numerator or the denominator-plus files. Weights should not be used when computing the total number of live births, or the number of live births by characteristics from the denominator-plus file, as the use of weights under these circumstances will yield a slight overestimate of the total number of US births. For multivariate analysis, the use of weights is generally recommended, however, a decision should be made on an individual basis, depending on the type of multivariate technique used, and the goals of the particular analysis.

Imputed birthweight

An imputation for not-stated birthweight has been added to the data set, to reduce potential bias in the computation of birthweight-specific infant mortality rates. Basically, if birthweight is not-

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

stated and the period of gestation is known, birthweight is assigned the value from the previous record with the same period of gestation, race, sex, and plurality. Imputed values are flagged. The addition of this imputation reduced the percent of not-stated responses for birthweight, thus reducing (but not eliminating) the potential for underestimation when computing birthweight-specific infant mortality rates.

Methodology

The methodology used to create the national file of linked birth and infant death records takes advantage of two existing data sources:

1. State linked files for the identification of linked birth and infant death certificates; and
2. NCHS natality and mortality computerized statistical files, the source of computer records for the two linked certificates.

Virtually all States routinely link infant death certificates to their corresponding birth certificates for legal and statistical purposes. When the birth and death of an infant occur in different States, copies of the records are exchanged by the State of death and State of birth in order to effect a link. In addition, if a third State is identified as the State of residence at the time of birth or death, that State is also sent a copy of the appropriate certificate by the State where the birth or death occurred.

The NCHS natality and mortality files, produced annually, include statistical data from birth and death certificates that are provided to NCHS by States under the Vital Statistics Cooperative Program (VSCP). The data have been coded according to uniform coding specifications, have passed rigid quality control standards, have been edited and reviewed, and are the basis for official U.S. birth and death statistics.

To initiate processing, NCHS obtained matching birth certificate numbers from States for all infant deaths that occurred in their jurisdiction. We used this information to extract final, edited mortality and natality data from the NCHS natality and mortality statistical files. Individual birth and death records were selected from their respective files and linked into a single statistical record, thereby establishing a national linked record file.

After the initial linkage, NCHS returned to the States where the death occurred computer lists of unlinked infant death certificates for follow up linking. If the birth occurred in a State different from the State of death, the State of birth identified on the death certificate was contacted to obtain the linking birth certificate. State additions and corrections were incorporated, and a final, national linked file was produced. Characteristics of the natality and mortality data from which the linked file is constructed are described in detail in the Technical Appendices and Addenda included in this document.

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

Characteristics of Unlinked File

For the 2000 birth cohort linked file 358, or 1.3% of all infant death records could not be linked to their corresponding birth certificates. Unlinked records are included in a separate data file in this data set. The unlinked record file uses the same record layout as the numerator file of linked birth and infant death records. However, except as noted below, tape locations 1-210, reserved for information from the matching birth certificate, are blank since no matching birth certificate could be found for these records. The sex field (tape location 79) contains the sex of infant as reported on the death certificate, rather than the sex of infant from the birth certificate, which is not available. The race field (tape location 36-37) contains the race of the decedent as reported on the death certificate rather than the race of mother as reported on the birth certificate as is the case with the linked record file. The race of mother on the birth certificate is generally considered to be more accurate than the race information from the death certificate (see section on *Comparison of race data from birth and death certificates* in the Mortality Technical Appendix included in this documentation). Also, date of birth as reported on the death certificate is used to generate age at death. This information is used in place of date of birth from the birth certificate, which is not available.

Documentation table 6 shows counts of unlinked records by race and age at death for each State of residence. The user is cautioned in using table 6 that the race and residence items are based on information reported on the death certificate; whereas, tables 1-5 present data from the linked file in which the race and residence items are based on information reported on the birth certificate. (see section on *Comparison of race data from birth and death certificates* in the Mortality Technical Appendix included in this documentation).

Percent of Records Linked

The 2000 birth cohort linked file includes 27,719 linked infant death records and 360 unlinked infant death records by place of occurrence. The linked file is weighted to the sum of linked plus unlinked records, thus the total number of weighted infant deaths by place of occurrence is 28,079. While the overall percent linked for infant deaths in the 2000 birth cohort linked file is 98.7%, there are differences in percent linked by certain variables. These differences have important implications for how the data is analyzed.

Table 1 shows the percent of infant deaths linked by State of residence. While most States link a high percentage of infant deaths, linkage rates for some States are well below the national average. When a high percentage of deaths remain unlinked, infant mortality rates computed for these States are underestimated. It is for this reason that weights were added to the linked files beginning with 1995 data, to correct for biases in the data due to poor data linkage for particular states.

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

Geographic classification

Geographic codes in this data set have been updated to reflect the results of the 1990 census, and differ slightly from those used in previous linked files. Because of confidentiality concerns, only those counties and cities with a population size of 250,000 or more are separately identified in this data set. Users should refer to the geographic code outline in this document for the list of available areas and codes.

For events to be included in the linked file, both the birth and death must occur inside the 50 States and D.C. in the case of the 50 States and D.C. file; or in Puerto Rico, the Virgin Islands or Guam in the case of the Puerto Rico, Virgin Islands and Guam file. In tabulations of linked data and denominator data events occurring in each of the respective areas to nonresidents are included in tabulations that are by place of occurrence, and excluded from tabulations by place of residence. These exclusions are based on the usual place of residence of the mother. This item is contained in both the denominator file and the birth section of the numerator (linked) file. Nonresidents are identified by a code 4 in location 11 of these files.

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

Table 1. Percent of infant deaths linked by state of residence of birth: United States, 2000 birth cohort

United States	98.7	Nebraska	99.4
Alabama	100	Nevada	96.6
Alaska	100	New Hampshire	98.8
Arizona	99.1	New Jersey	96.7
Arkansas	99.7	New Mexico	93.8
California	98.1	New York State (no NYC)	98.4
Colorado	100	New York City	99.9
Connecticut	99.6	North Carolina	99.5
Delaware	99.0	North Dakota	100
District of Columbia	98.9	Ohio	95.9
Florida	99.9	Oklahoma	92.8
Georgia	100	Oregon	100
Hawaii	97.2	Pennsylvania	99.7
Idaho	100	Rhode Island	100
Illinois	99.2	South Carolina	100
Indiana	98.5	South Dakota	100
Iowa	100	Tennessee	100
Kansas	98.2	Texas	97.0
Kentucky	99.2	Utah	98.5
Louisiana	97.0	Vermont	100
Maine	98.5	Virginia	98.9
Maryland	98.3	Washington	99.8
Massachusetts	99.2	West Virginia	100
Michigan	99.9	Wisconsin	100
Minnesota	99.7	Wyoming	100
Mississippi	100		
Missouri	99.5		
Montana	100		

Demographic and Medical Classification

The documents listed below describe in detail the procedures employed for demographic classification on both the birth and death records and medical classification on death records. These documents, while not absolutely essential to the proper interpretation of the data for a number of general applications, should nevertheless be studied carefully prior to any detailed analysis of demographic or medical data variables. In particular, there are a number of exceptions to the ICD rules in multiple cause-of-death coding which, if not treated properly, may result in faulty analysis of the data. Volumes 1, 2 and 3 of the ICD-10 may be purchased from the World Health Organization (WHO) Publication Center USA, 49 Sheridan Avenue, Albany,

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

New York, 12210 (<http://www.who.int/whosis/icd10/index.html>). Many of the instruction manuals listed below are available electronically on the NCHS website at: <http://www.cdc.gov/nchs/about/major/dvs/im.htm>. In addition, users who do not already have access to these documents may request them from the Chief, Mortality Medical Classification Branch, Division of Vital Statistics, National Center for Health Statistics, 4105 Hopson Road, Research Triangle Park, North Carolina 27709. The technical appendices for natality and mortality included in this document also provide information on the source of data, coding procedures, quality of the data, etc.

- A. National Center for Health Statistics. Vital statistics, Instructions for Classifying the Underlying Cause-of-Death, 2000. NCHS Instruction Manual, Part 2a. Hyattsville, Maryland: Public Health Service.
- B. National Center for Health Statistics. Vital statistics, Instructions for Classifying Multiple Cause-of-Death, 2000. NCHS Instruction Manual, Part 2b. Hyattsville, Maryland: Public Health Service.
- C. National Center for Health Statistics. Vital statistics, ICD-10 ACME Decision Tables for Classifying Underlying Causes-of-Death, 2000. NCHS Instruction Manual, Part 2c. Hyattsville, Maryland: Public Health Service.
- D. National Center for Health Statistics. Vital statistics, NCHS Procedures for Mortality Medical Data System File Preparation and Maintenance, Effective 2000. NCHS Instruction Manual, Part 2d. Hyattsville, Maryland: Public Health Service.
- E. National Center for Health Statistics. Vital statistics, ICD-10 TRANSAX Disease Reference Tables for Classifying Multiple Causes-of-Death, 1999. NCHS Instruction Manual, Part 2f. Hyattsville, Maryland: Public Health Service.
- F. National Center for Health Statistics. Vital statistics, Classification and Coding Instructions for Live Birth Records, 1999. NCHS Instruction Manual, Part 3a. Hyattsville, Maryland: Public Health Service.

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

- G. National Center for Health Statistics. Vital statistics, Demographic Classification and Coding Instructions for Death Records, 2000. NCHS Instruction Manual, Part 4. Hyattsville, Maryland: Public Health Service.
- H. National Center for Health Statistics. Vital statistics, Computer Edits for Natality Data, Effective 1993. NCHS Instruction Manual Part 12. Hyattsville, Maryland: Public Health Service.
- I. National Center for Health Statistics. Vital statistics, Computer Edits for Mortality Data, Effective 1999. NCHS Instruction Manual Part 11. Hyattsville, Maryland: Public Health Service.

Change in Cause-of-Death Classification

In data year 1999, a new classification system for coding causes of death was implemented in the United States: the Tenth Revision of the International Classification of Diseases (ICD-10), developed by the World Health Organization (WHO). Information about the new system can be obtained at the following address: <http://www.cdc.gov/nchs/about/major/dvs/icd10des.htm>

Underlying Cause of Death Data

Mortality statistics by cause of death are compiled from entries on the medical certification portion of the death certificate. The U.S. Standard Certificate of Death is shown in the Mortality Technical Appendix which is included in this documentation. Causes of death include “all those diseases, morbid conditions or injuries which either resulted in or contributed to death and the circumstances of the accident or violence which produced these injuries”. The medical certification of death is divided into two sections. In Part I, the physician is asked to provide the causal chain of morbid conditions that led to death, beginning with the condition most proximate to death on line (a) and working backwards to the initiating condition. The lines (a) through (d) in Part I are connected by the phrase “due to, or as a consequence of.” They were designed to encourage the physician to provide the causally related sequence of medical conditions that resulted in death. Thus, the condition on line (a) should be due to the condition on line (b), and the condition on line (b) should be a consequence of the condition on line (c), etc., until the full sequence is described back to the originating or initiating condition. If only one step in the chain of morbid events is recorded, a single entry on line (a) is adequate. Part I of the medical certification is designed to facilitate the selection of the underlying cause of death when two or more causes are recorded on the certificate. The underlying cause of death is defined by the WHO in the ICD-10 as “(a) the disease or injury which initiated the chain of morbid events leading directly to death, or (b) the circumstances of the accident or violence that produced the fatal injury” and is generally considered the most useful cause from a public health standpoint. Part II of the cause-of-death section of the death certificate solicits other conditions that the certifier believed contributed to death, but were not in the causal chain. While some details of

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

the death certificate vary by State, all States use the same general format for medical certification outlined in the U.S. Standard Certificate. The U.S. Standard Certificate, in turn, closely follows the format recommended by the WHO.

If the death certificate is properly completed, the disease or condition listed on the lowest used line in Part I is usually accepted as the underlying cause of death. This is an application of “The General Principle.” The General Principle is applied unless it is highly improbable that the condition on the lowest line used could have given rise to all of the diseases or conditions listed above it. In some cases, the sequence of morbid events entered on the death certificate is not specified correctly. A variety of errors may occur in completing the medical certification of death. Common problems include the following: The causal chain may be listed in reverse order; the distinction between Part I and Part II may have been ignored so that the causal sequence in Part I is simply extended unbroken into Part II; or the reported underlying cause is unlikely, in an etiological sense, to have caused the condition listed above it. In addition, sometimes the certifier attributes the death to uninformative causes such as cardiac arrest or pulmonary arrest.

To resolve the problems of incorrect or implausible cause-of-death statements, the WHO designed standardized rules to select an underlying cause of death from the information available on the death certificate that is most informative from a public health perspective. The rules for the Tenth Revision as updated by WHO since the publication of ICD-10 are described in NCHS instruction manual Part 2A. Coding rules beyond the General Principle are invoked if the cause-of-death section is completed incorrectly or if their application can improve the specificity and characterization of the cause of death in a manner consistent with the ICD. The rules are applied in two steps: selection of a tentative underlying cause of death, and modification of the tentative underlying cause in view of the other conditions reported on the certificate in either Part I or Part II. Modification involves several considerations by the medical coder: determining whether conditions in Part II could have given rise to the underlying cause, giving preference to specific terms over generalized terms, and creating linkages of conditions that are consistent with the terminology of the ICD.

For a given death, the underlying cause is selected from the condition or conditions recorded by the certifier in the cause-of-death section of the death certificate. NCHS is bound by international agreement to make the selection of the underlying cause through the use of the ICD-10 classification structure, and the selection and modification rules contained in this revision of the ICD. These rules are contained in a computer software program called ACME (Automated Classification of Medical Entities). ACME does exactly what a coder would do to select the underlying cause of death. The ACME program has been used for final mortality data since 1968.

The WHO selection rules take into account the certifier’s ordering of conditions and their causal relationships to systematically identify the underlying cause of death. The intent of these rules is

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

to improve the usefulness of mortality statistics by giving preference to certain classification categories over others and consolidating two or more conditions on the certificate into a single classification category.

Multiple Cause of Death Data

The limitations of the underlying cause concept and the need for more comprehensive data suggested the need for coding and tabulating all conditions listed on the death certificate. Coding all listed conditions on the death certificate was designed with two objectives in mind. First, to facilitate studies of the relationships among conditions reported on the death certificate, which require presenting each condition and its location on the death certificate in the exact manner given by the certifier. Secondly, the coding needed to be carried out in a manner by which the underlying cause-of-death could be assigned using the WHO coding rules. Thus, the approach in developing multiple cause data was to provide two fields: 1) entity axis and 2) record axis. For entity axis, NCHS suspends the provisions of the ICD that create linkages between conditions for the purpose of coding each individual condition, or entity, with minimum regard to other conditions present on the death certificate.

Record axis is designed for the generation of person-based multiple cause statistics. Person-based analysis requires that each condition be coded within the context of every other condition on the same death certificate and modified or linked to such conditions as provided by ICD-10. By definition, the entity data cannot meet this requirement since the linkage provisions modify the character and placement of the information originally recorded by the certifier. Essentially, the axis of the classification has been converted from a entity basis to a record (or person) basis. The record axis codes are assigned in terms of the set of codes that best describe the overall medical certification portion of the death certificate.

This translation is accomplished by a computer system called TRANSAX (Translation of Axis). TRANSAX selectively uses the traditional linkage and modification rules for mortality coding. Underlying cause linkages which simply prefer one code over another for purposes of underlying cause selection are not included. Each entity code on the record is examined and modified or deleted as necessary to create a set of codes that are free of contradictions and are the most precise within the constraints of ICD-10 and medical information on the record. Repetitive codes are deleted. The process may 1) combine two entity axis categories together to a new category thereby eliminating a contradiction or standardizing the data; or 2) eliminate one category in favor of another to promote specificity of the data or resolve contradictions. The following examples from ICD-10 illustrate the effect of this translation:

Case 1: When reported on the same record as separate entities, cirrhosis of liver and alcoholism are coded to K74.6 (Other and unspecified cirrhosis of liver) and F10.2 (Mental and behavioral disorders due to use of alcohol; dependence

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

syndrome), respectively. Tabulation of records with K74.6 would imply that such records had no mention of alcohol. A preferable code would be K70.3 (Alcoholic cirrhosis of liver) in lieu of both K74.6 and F10.2.

Case 2: If “gastric ulcer” and “bleeding gastric ulcer” are reported on a record they are coded to K25.9 (Gastric ulcer, unspecified as acute or chronic, without mention of hemorrhage or perforation) and K25.4 (Gastric ulcer, chronic or unspecified with hemorrhage), respectively. A more concise code is K25.4 which shows both the gastric ulcer and the bleeding.

Entity Axis Codes

The original conditions coded for selection of the underlying cause-of-death are reformatted and edited prior to creating the public-use data file. The following paragraphs describe the format and application of entity axis data.

1. Format. Each entity-axis code is displayed as an overall seven byte code with subcomponents as follows:

1. Line indicator: The first byte represents the line of the death certificate on which the code appears. Six lines (1-6) are allowable with the fourth and fifth denoting one or two written in “due to”s beyond the three lines provided in Part I of the U.S. standard death certificate. Line “6” represents Part II of the death certificate.
2. Position indicator: The next byte indicates the position of the code on the line, i.e., it is the first (1), second (2), third (3) eighth (8) code on the line.
3. Cause category: The next four bytes represent the ICD-10 cause code.
4. The last byte is blank.

A maximum of 20 of these seven byte codes are captured on a record for multiple cause purposes. This may consist of a maximum of 8 codes on any given line with up to 20 codes distributed across three or more lines depending on where the subject conditions are located on the certificate. Codes may be omitted from one or more lines, e.g., line 1 with one or more codes, line 2 with no codes, line 3 with one or more codes.

In writing out these codes, they are ordered as follows: line 1 first code, line 1 second code, etc. -
---- line 2 first code, line 2 second code, etc. ----- line 3 ---- line 4 ----- line 5 ----- line 6. Any space remaining in the field is left blank. The specifics of locations are contained in the record layout given later in this document.

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

2. Edit. The original conditions are edited to remove invalid codes, reverify the coding of certain rare causes of death, and assure age/cause and sex/cause compatibility. Detailed information relating to the edit criteria and the sets of cause codes which are valid to underlying cause coding and multiple cause coding are provided in NCHS Instruction Manual Part 11.

3. Entity Axis Applications. The entity axis multiple cause data file is appropriate for analyses that require that each condition be coded as a stand alone entity without linkage to other conditions and/or require information on the placement of such conditions in the death certificate. Within this framework, the entity data are appropriate to examine relationships among conditions and the validity of traditional assumptions in underlying cause selection. Additionally, the entity data provide in certain categories a more detailed code assignment that could be excluded in creating record axis data. Where such detail is needed for a study, the user should use entity data. Finally, the researcher may not wish to be bound by the assumptions used in the axis translation process.

The main limitation of entity axis data is that it does not necessarily reflect the best code for a condition when considered within the context of the medical certification as a whole. As a result, certain entity codes can be misleading or even contradict other codes in the record. For example, category K80.2 is titled “Calculus of gallbladder without cholecystitis.” Within the framework of entity codes this is interpreted to mean that the codable entity itself contained no mention of cholecystitis rather than that cholecystitis was not mentioned anywhere on the record. Tabulation of records with a “K80.2” as a count of persons having Calculus of gallbladder without cholecystitis would therefore be erroneous. This illustrates the fact that under entity coding the ICD-10 titles cannot be taken literally. The user should study the rules for entity coding as they relate to his/her research prior to use of entity data. The user is further cautioned that the inclusion notes in ICD-10 that relate to modifying and combining categories are seldom applicable to entity coding (except where provided NCHS Instruction Manual Part 2b).

In tabulating the entity axis data, one may count codes with an individual code representing the number of times the condition(s) appears in the file. In this kind of tabulation of morbid conditions, the counts among categories may be added together to produce counts for groups of codes. Alternatively, subject to the limitations given above, one may count persons having mention of the disease represented by a code or codes. In this instance it is not correct to add counts for individual codes to create person counts for groups of codes. Since more than one code in the researcher’s interest may appear together on the certificate, totaling must account for higher order interactions among codes. Up to 20 codes may be assigned on a record; therefore, a 20-way interaction is theoretically possible. All totaling must be based on mention of one or more of the categories under investigation.

Record Axis Codes

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

The following paragraphs describe the format and application of record-axis data. Part 2f of the Instruction Manual Series (ICD-10 TRANSAX Disease Reference Tables for classifying Multiple Causes-of-Death) describes the TRANSAX process for creating record axis data from entity axis data.

1. Format. Each record (or person) axis code is displayed in five bytes. Location information is not relevant. The Code consists of the following components:

1. Cause category: The first four bytes represent the ICD-10 cause code.
2. The last byte is blank.

Again, a maximum of 20 codes are captured on a record for multiple cause purposes. The codes are written in a 100-byte field in ascending code number (5 bytes) order with any unused bytes left blank.

2. Edit. The record axis codes are edited for rare causes and age/cause and sex/cause compatibility. Likewise, individual code validity is checked. The valid code set for record axis coding is the same as that for entity coding.

3. Record Axis Applications. The record axis multiple cause data are the basis for NCHS core multiple cause tabulations. Location of codes is not relevant to this data, and conditions have been linked into the most meaningful categories for the certification. The most immediate consequence for the user is that the codes on the record already represent mention of a disease assignable to that particular ICD-10 category. This is in contrast to the entity code which is assigned each time such a disease is reported on different lines of the certification. Secondly, the linkage implies that within the constraints of ICD-10 the most meaningful code has been assigned. The translation process creates for the user a data file that is edited for contradictions, duplicate codes, and imprecisions. In contrast to entity axis data, record axis data are classified in a manner comparable to underlying cause of death classification thereby facilitating joint analysis of these variables. A potential disadvantage of record axis data is that some detail is sacrificed in a number of the linkages.

The user can take the record axis codes as literally representing the information conveyed in ICD-10 category titles. While knowledge of the rules for combining and linking and coding conditions is useful, it is not a prerequisite to meaningful analysis of the data as long as one is willing to accept the assumptions of the axis translation process. The user is cautioned, however, that due to special rules in mortality coding, not all linkage notes in ICD-10 are used. (NCHS Instruction Manual Part 2f).

The user should proceed with caution in using record axis data to count conditions as opposed to people with conditions, since linkages have been invoked and duplicate codes have been

Linked Birth/Infant Death Data Set — 2000 Birth Cohort

eliminated. As with entity data, person-based tabulations that combine individual cause categories must take into account the possible interaction of up to 20 codes on a single certificate.

Additional Information

In using the NCHS multiple cause data files, the user is urged to review the information in this document and its references. The instructional material does change from year to year and ICD revision to ICD revision. The user is cautioned that coding of specific ICD-10 categories should be checked in the appropriate instruction manual. What may appear on the surface to be the correct code by ICD-10 may in fact not be correct as given in the instruction manuals.

If on the surface it is not obvious whether entity axis or record axis data should be employed in a given application, detailed examination of NCHS Instruction Manual Part 2f and its attachments will probably provide the necessary information to make a decision. It allows the user to determine the extent of the trade-offs between the two sets of data in terms of specific categories and the assumption of axis translation. In certain situations, a combination of entity and record axis data may be the more appropriate alternative.